Fake News Detection

Karolina Mączka, Tymoteusz Urban

DATA EXPLORATION & PREPROCESSING
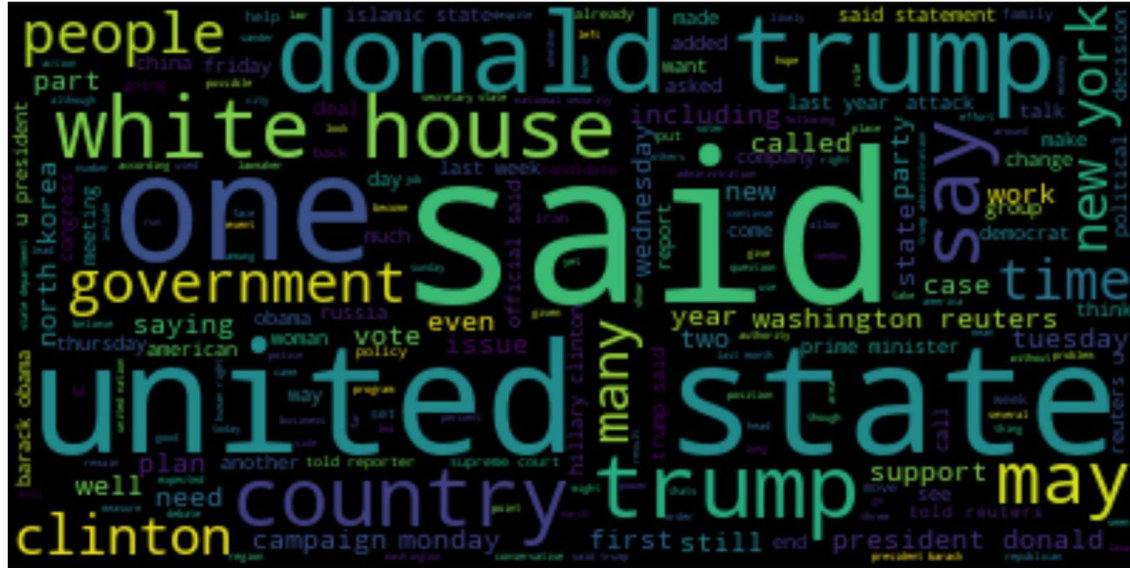
CHOOSING CLASSIFIER

Classifier	precision	recall	f1 score	auc score	log loss
Logistic regression	0.956505	0.973144	0.961124	0.987174	0.162712
Multinomial Naive Bayes	0.924676	0.962241	0.898798	0.938696	0.342787
SVM	0.945115	0.967933	0.940562	0.913715	NaN
SVM with bigrams	0.951745	0.973625	0.929085	0.891667	NaN
SVM with huber	0.951013	0.970338	0.952733	0.933322	NaN
XG boost	0.975960	0.985249	0.971234	0.994027	0.094381
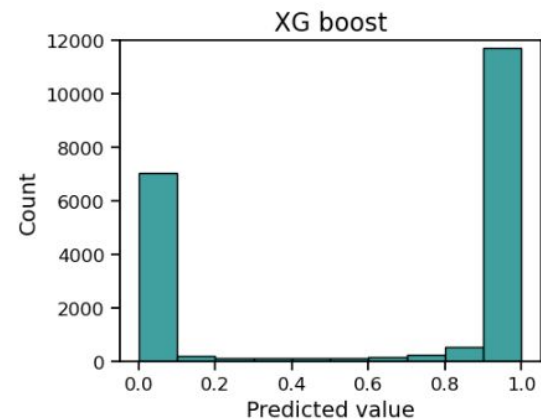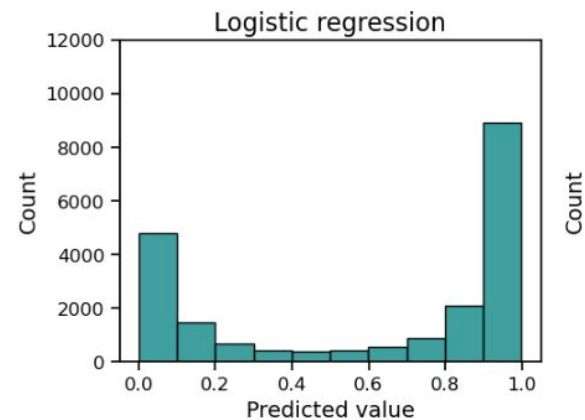
OPTIMIZING MODEL

Classifier	precision	recall	f1 score	auc score	log loss
XG boost	0.975960	0.985249	0.971234	0.994027	0.094381
XGB no tfidf	0.975976	0.985249	0.971426	0.993567	0.095417
XGB with bigrams	0.975277	0.984768	0.971951	0.994031	0.092696

Classifier	precision	recall	f1 score	auc score	log loss
XG boost	0.975960	0.985249	0.971234	0.994027	0.094381
XGB boost final	0.974167	0.984207	0.969135	0.994371	0.096173

FINAL MODEL

XG boost ROC curve

AUC = 99.4027